Is Generative AI a Game-Changer for Computational Materials Discovery of New Solid-State Materials?

George E.H. Smith[1], Abigail Parsons[1], P. M. Anuththara Bandaranayake[1], Peter J. Graham[1], Alexander G. Squires[1], David O. Scanlon[1]

[1]School of Chemistry, University of Birmingham, Birmingham, UK

ges397@student.bham.ac.uk

The computational discovery of novel materials for technological applications remains a challenging and highly sought after goal in computational materials chemistry. Recent advances in artificial intelligence (AI) and the development of large language models (LLMs) have transformed the approaches to crystal structure prediction (CSP) problems and materials discovery. AI models have been developed to overcome the high computational cost associated with using traditional methods such as density functional theory (DFT) for CSP and touted the discovery of many previously unknown materials[1]. However, crystal structure prediction relies on the reliable identification of the ground state structure, and it remains unclear how well current AI driven methods can reproduce true ground state structures without explicit first principles validation.

In this work, we present a systematic comparison between three LLM-based generative crystal structure prediction frameworks, MatterGen[2], CrystaLLM[3], and Chemeleon[4], and the established *ab initio* random structure searching (AIRSS)[5] method. A diverse and well-studied set of binary materials - NaCl, ZnS, CuS, MoS_2, Sb_2Se_3, AgO, and Ta_2O_5 were used to evaluate the efficacy of each method for crystal structure prediction by evaluating the ability to identify experimentally known ground states and known metastable states of each compound. Additionally, three out-of-distribution cases were tested, $CaZrCl_6$, Li_2NH and $NaTaOCl_4$ to test the performance of LLM models on compositions where limited data is available and the ability to predict low energy structure trends in more complex materials.

We find that AIRSS remains reliable for exhaustive ground state searches, while the AI based models rapidly generate chemically plausible and structurally diverse candidate structures. However, the generative models alone do not reliably predict the lowest energy polymorphs and frequently misidentify metastable structures in higher frequencies. Overall, our results indicate that current AI driven crystal structure prediction methods are best used as efficient structure generation tools within hybrid AI and DFT workflows, rather than as standalone predictors of thermodynamic stability. This work provides a clear assessment of where generative AI adds value in crystal structure prediction and where first principles methods remain essential.

References

[1] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon and E. D. Cubuk, *Nature*, 2023, **624**, 80–85.

[2] C. Zeni, R. Pinsler, D. Zügner, A. Fowler, M. Horton, X. Fu, Z. Wang, A. Shysheya, J. Crabbé, S. Ueda, R. Sordillo, L. Sun, J. Smith, B. Nguyen, H. Schulz, S. Lewis, C.-W. Huang, Z. Lu, Y. Zhou, H. Yang, H. Hao, J. Li, C. Yang, W. Li, R. Tomioka and T. Xie, *Nature*, 2025, **639**, 624–632.

[3] L. M. Antunes, K. T. Butler and R. Grau-Crespo, *Nat Commun*, 2024, **15**, 10570.

[4] H. Park, A. Onwuli and A. Walsh, *Nat Commun*, 2025, **16**, 4379.

[5] C. J. Pickard and R. J. Needs, *J. Phys.: Condens. Matter*, 2011, **23**, 053201.

Figures

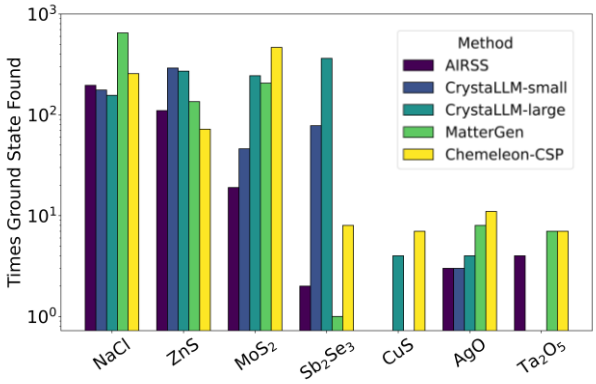

Figure 1. Times ground state crystal structure found with each structure generation method out of 1000 generated structures.